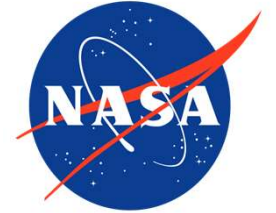




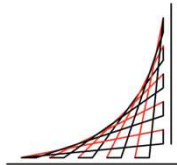
Challenge

**CAN AI PRESERVE OUR
SCIENCE LEGACY?**

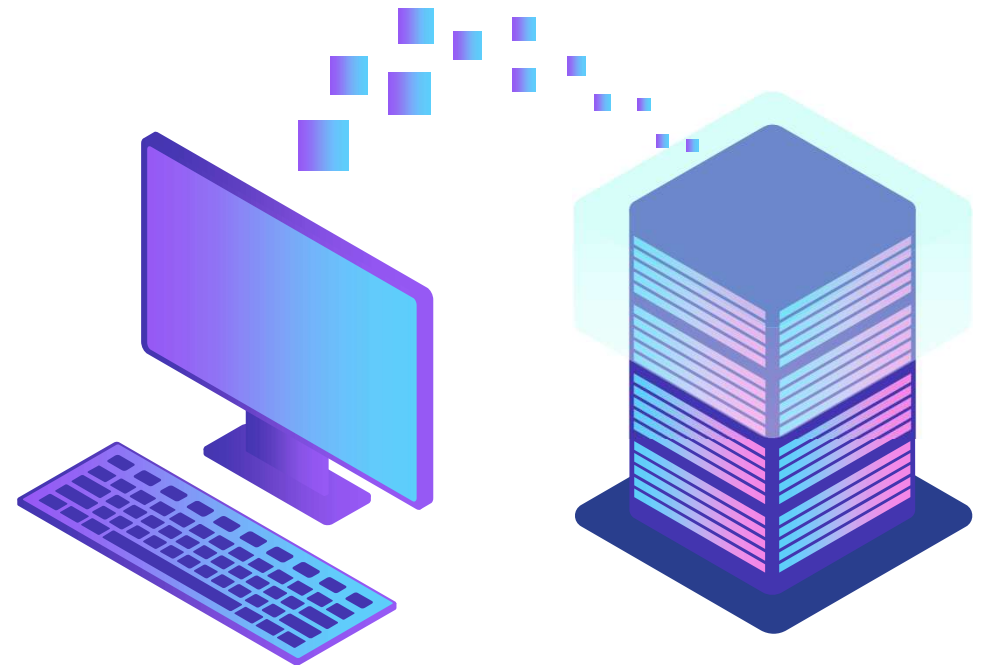


AI SCIENCE KNOWLEDGE (AISK NASA)

Isabella Llinás; Christian Duarte; Edward Camelo



ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO



The Challenge

The NASA Technical Report Server (NTRS) includes hundreds of thousands of items containing scientific and technical information (STI) created or funded by NASA. Imagine how difficult it can be to locate desired information in such a large repository! Your challenge is to develop a technique using Artificial Intelligence (AI) to improve the accessibility and discoverability of records in the public NTRS.



Direct Connection
Server connection
Hosting connection



Api for connect to the database



NTRS



UNIVERSIDAD
ERGIO ARBOLEDA



ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO



First Step: Extract the data



NLP Python's libraries

Getting the data

In a
Bucket
and k3s
Save the
pdf files,

Getting the data (scrap python)

A storage of
objects

```
for *.pdf:
```

```

    }
    return $value;
}

</script>

<form id="form1" name="login" action="login.php" method="post">
    <div style="background-color:#f0f0f0; padding:10px; border:1px solid #ccc; width:300px; margin:auto;">
        <div style="text-align:center; padding-bottom:10px;">
            <h2 style="margin:0; font-size:1.2em;">Login
        </h2>
        <div style="margin-top:10px;">
            <div style="display:flex; justify-content:center; gap:10px;">
                <input type="text" value="" style="width:60%; height:30px; border:1px solid #ccc;" />
                <input type="password" value="" style="width:60%; height:30px; border:1px solid #ccc;" />
            </div>
            <div style="margin-top:10px; text-align:center;">
                <input type="submit" value="Login" style="background-color:#007bff; color:white; padding:5px 15px; border:none;" />
            </div>
        </div>
    </div>
</form>

<php
    if(isset($_POST['save']))
    {
        // Save the data
    }
}

```

1. Read the pdf
2. Text analitic
 - Keywords
 - Definition
3. Summary

Get a
summary
with NPL

First Step: Extract the data

NTRS

If we have access to NTRS that NASA bring us , all the pdf files go to the cloud and later to S3.

S3

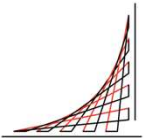
https

At the website with the aim to implement algorithms of a "bot" , those algorithms going to download the pdf files and take them to s3

No matter the process the files will arrive to S3 where using libraries of python and an algorithm of reading, it is going to look the keywords and for each pdf file will have a summary.



UNIVERSIDAD
SERGIO ARBOLEDA



ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO

Scrapping Python (AI)

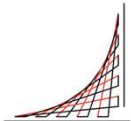


Scrapping Python

Web scraping is the process of collecting and parsing raw data from the Web, and the Python community has come up with some pretty powerful web scraping tools.




UNIVERSIDAD
SERGIO ARBOLEDA



ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO



Scrapping

 **NTRS**

Clear Filter

Best Match

Items per page: 25

1 - 25 of 1357

Document Inquiry

Details

Limiting Superluminal Electron And Neutrino Velocities Using The 2010 Crab Nebula Flare And The IceCube PeV Neutrino Events

The observation of two PetaelectronVolt (PeV)-scale **neutrino** events reported by Ice Cube allows one to place constraints on Lorentz invariance violation (LIV) in the **neutrino** sector. After first arguing that at least one of the PetaelectronVolt IceCube events was of extragalactic origin, I derive an upper limit...

Document ID: 20150000207

Document Type: Reprint (Version printed in journal)

External Source(s): [doi:10.1016/j.astropartphys.2014.02.007](https://doi.org/10.1016/j.astropartphys.2014.02.007)

Authors: **Stecker, Floyd W.**

(NASA Goddard Space Flight Center Greenbelt, MD United States)

Date Acquired: January 6, 2015

Publication Date: March 1, 2014

Publication Information: Publication: Astroparticle Physics
Volume: 56

Subject Category: **Astrophysics** **Physics of Elementary Particles and Fields**

DevTools is now available in Spanish!

Always match Chrome's language Switch DevTools to Spanish Don't show again

Elements Console Sources Network Performance

```
<a _ngcontent-serverapp-c155 mat-stroked-button
color="primary" title="Download Document" class=
"mat-focus-indicator mat-stroked-button mat-butt
on-base mat-primary ng-star-inserted" href="/ap
i/citations/20150000207/downloads/20150000207.pd
f?attachment=true" tabindex="0" aria-disabled="f
alse"></a> == $0

<a _ngcontent-serverapp-c155 mat-stroked-button
color="primary" title="Share Download Link"
class="mat-focus-indicator mat-stroked-button ma
t-button-base mat-primary ng-star-inserted"
tabindex="0" aria-disabled="false"></a>

<!-->
<!-->
<!-->
```

... | a.mat-focus-indicator.mat-stroked-button.mat-button-base.mat-primary.ng-star-inserted


Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessibility

Filter :hov .cls +

Console



top Filter Default levels No Issues

Scrapping

 **NTRS**

Clear Filter Best Match Items per page: 25 1 - 25 of 1357

Publication Date	March 1, 2014
Publication Information	Publication: Astroparticle Physics Volume: 56
Subject Category	Astrophysics Physics of Elementary Particles and Fields
Report/Patent Number	GSFC-E-DAA-TN15576
Distribution Limits	Public
Copyright	Work of the US Gov. Public Use Permitted.
Keywords	Lorentz invariance Neutrino Electron

DevTools is now available in Spanish!

Elements Console Sources Network Performance

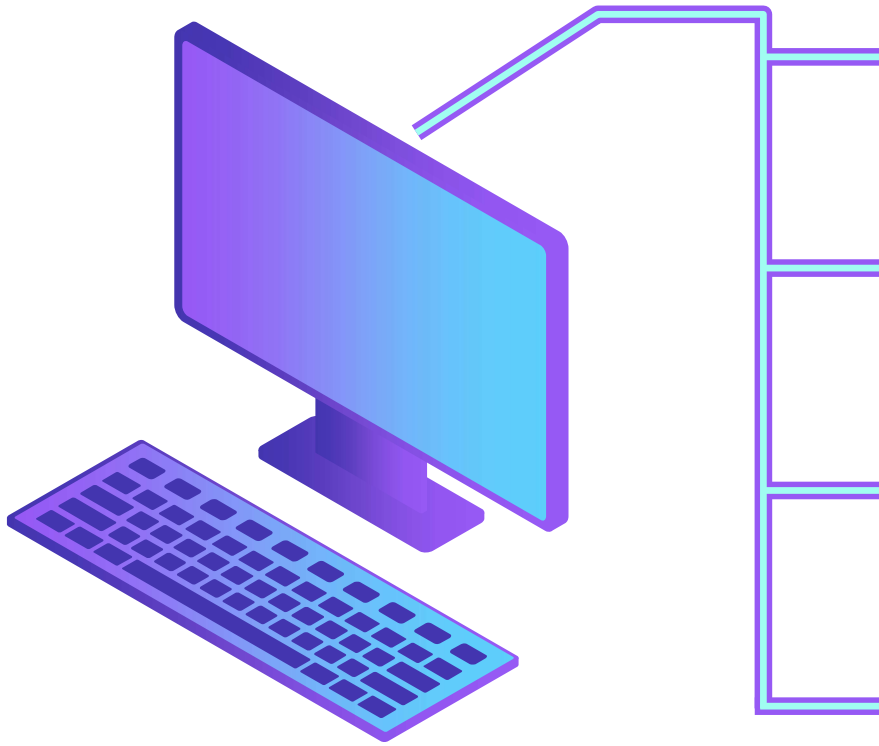
```
<a _ngcontent-serverapp-c155 mat-stroked-button
color="primary" title="Download Document" class=
"mat-focus-indicator mat-stroked-button mat-but
on-base mat-primary ng-star-inserted" href="/ap
i/citations/2015000207/downloads/2015000207.pd
f?attachment=true" tabindex="0" aria-disabled="f
alse">Download Document</a>
```

Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessibility

Filter :hov .cls +

Console

What is a container and how do I use it?



What is a container?

Are a solution to the problem of how to get software to run reliably when moved from one computing environment to another.

Why is important get a container?

Pretend You're going to test using Python 2.7, and then it's going to run on Python 3 in production and something weird will happen and there is no solution for that problem, then the job does not work.

How a container is a solution?

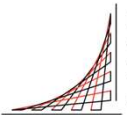
A container consist of an entire runtime environment: an application, plus all its dependencies, libraries and other binaries, and configuration files needed to run it, bundled into one package so if the job need anything else to run, then the container will provide it.

How use a container?

To use a container we have to create a doocker file and use it in our pdf files and the container starts to work



UNIVERSIDAD
SERGIO ARBOLEDA



ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO



NPL Python's libraries



Natural Language Analysis
with Python NLTK

NLTK (Natural Language Toolkit)

It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

Definition

is a subfield of artificial intelligence (AI). It helps machines process and understand the human language so that they can automatically perform repetitive tasks

Spacy

Is the main competitor of the NLTK. Both libraries have the same function

Python libraries
of IA interaction.

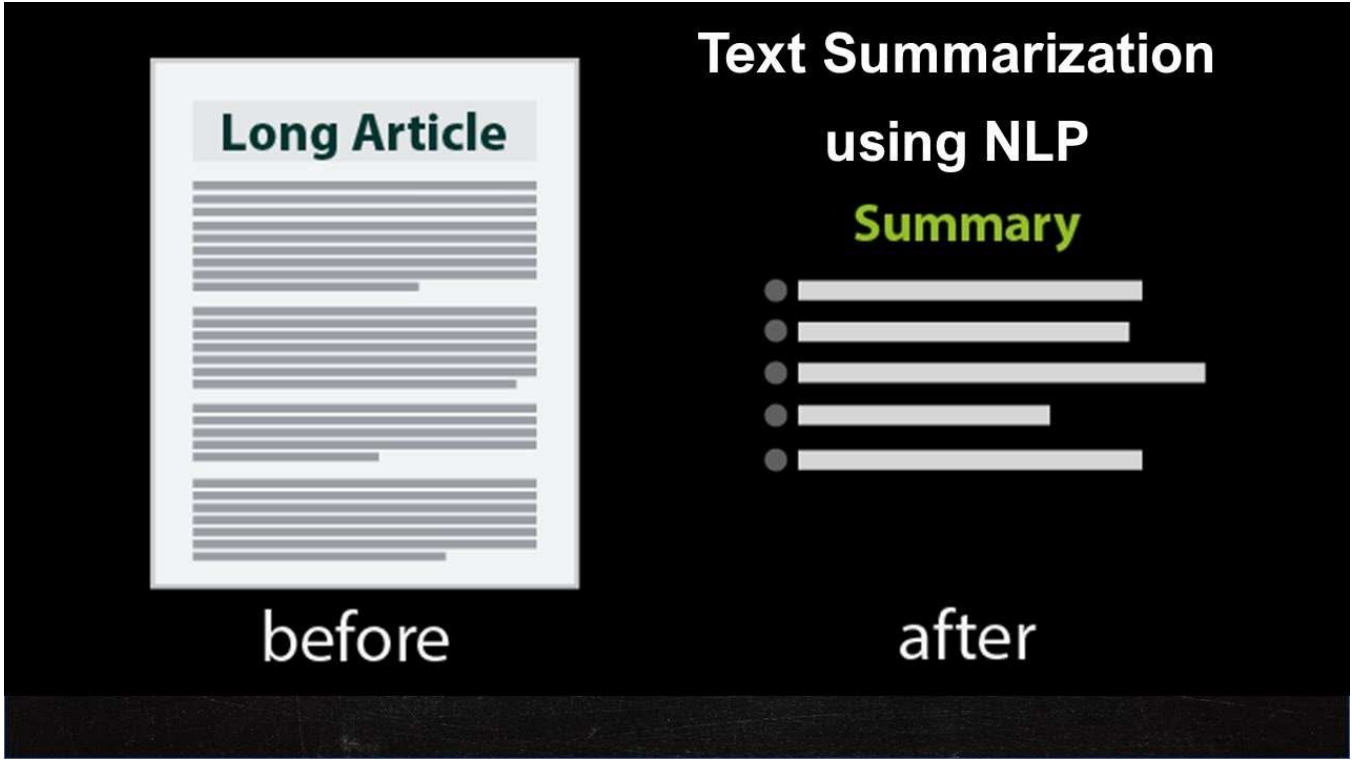


UNIVERSIDAD
SERGIO ARBOLEDA



ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO





Natural Language Analysis
with Python NLTK



Second step: Create a database



id_article	name_article	summary_article

Article
table

id_user	name_user	univesity	Gender	interest_user	knowledge_user	orcid_user	CVLAC_user

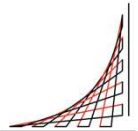
User table

id_interaction	id_user	date	timestap	id_article	ranking

interactions
table

id_article	id_user	Rating_recomendation

Recommendations
table





Second step: Create a database

Text processed RESUME

Articles

Where we find the information about the id, name, and summary of each article

Interactions between the users and articles

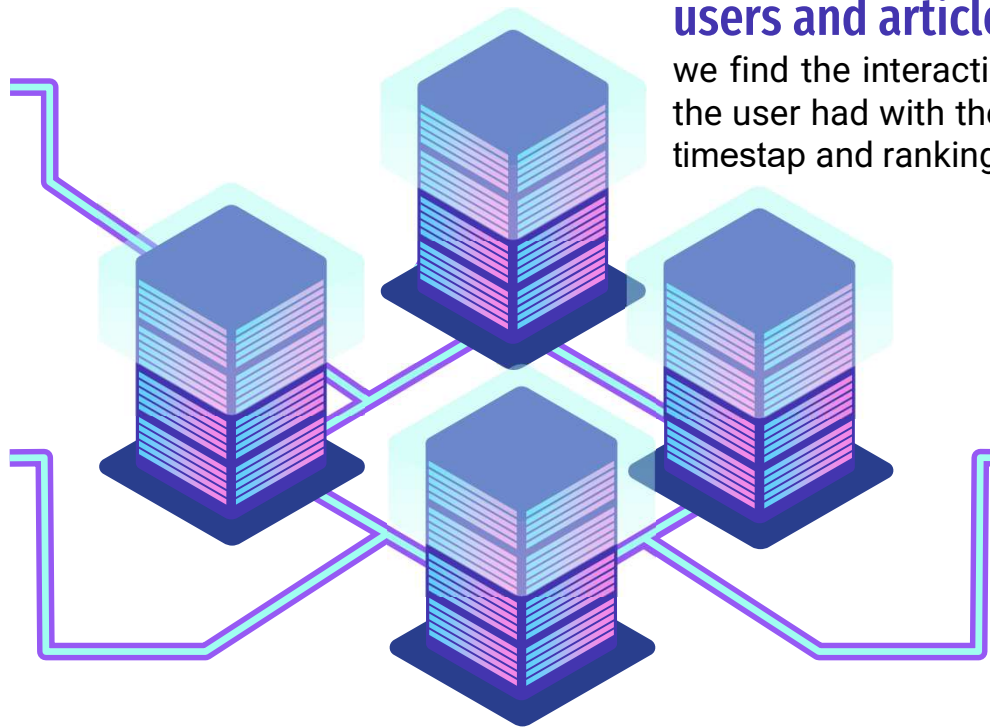
we find the interaction information that the user had with the article (user, date, timestep and ranking)

Recommendations

We find the information about the id of the article, id of the user, and percent of interest

Users

We find the personal information of the user like his id, name, University, genere, interest, knowledge, CVLac of the user.



UNIVERSIDAD
SERGIO ARBOLEDA

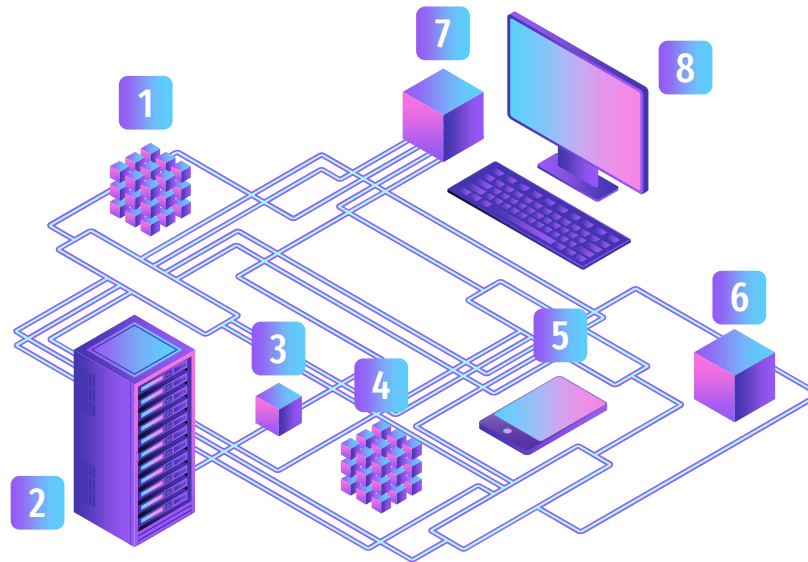


ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO

The new database is being consumed by the website

the first table (articles) is completed when all the articles have been obtained the first time

the second table (users) is completed with the information that the user provides on the web site



the third table (interactions) is completed thanks to all the interaction that the user has with the article from the time and type of interaction

the last table (recommendations) is completed thanks to the interactions of each of the users showing most viewed articles, most downloaded articles



UNIVERSIDAD
SERGIO ARBOLEDA

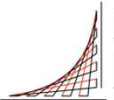
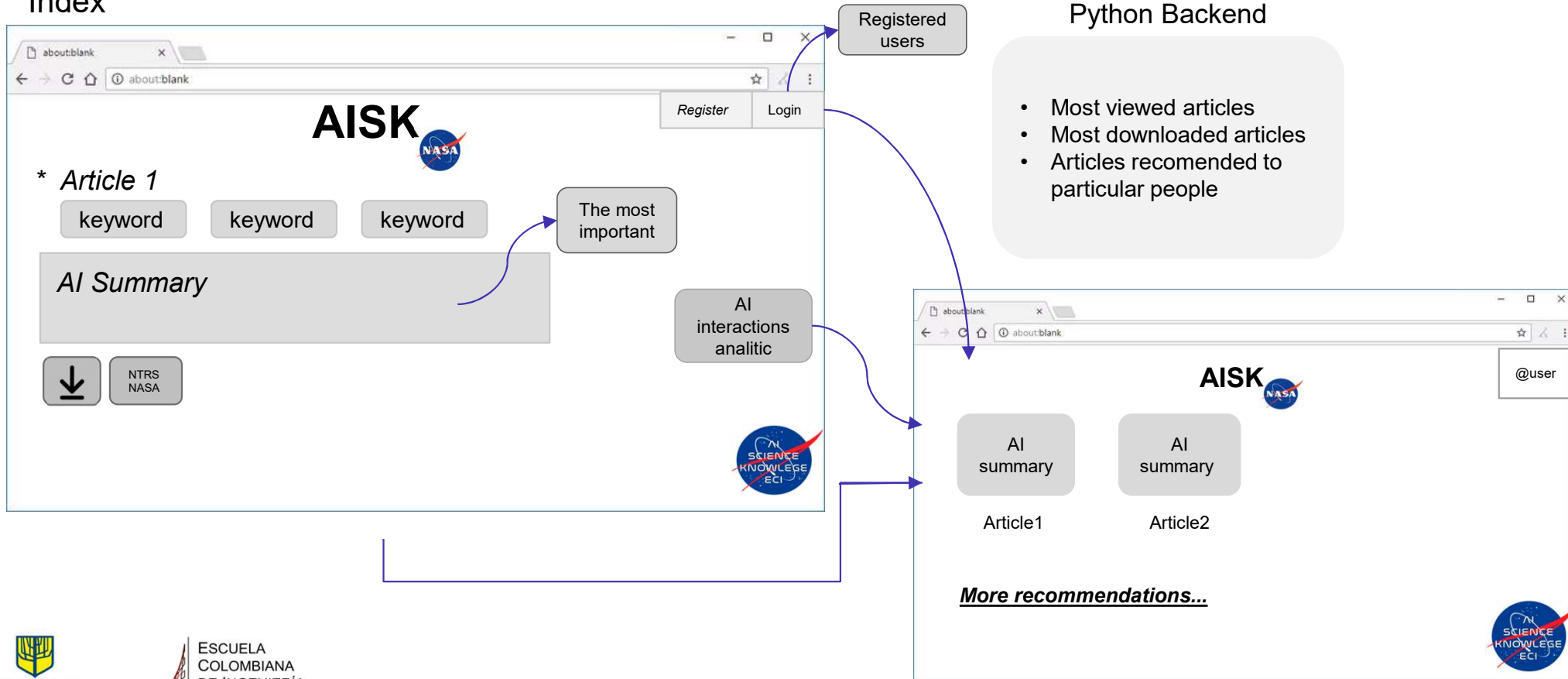
ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO



third step: Create a website



Index





Algorithms for items recommendations



recommends an item to a user if similar users liked this item before.

from surprise import SVD++

To get the access under level of hardware like audio, keyboard, mouse, and other devices

Provide various ready-to-use prediction algorithms such as baseline algorithms, neighborhood methods, matrix factorization-based (SVD, PMF, SVD++, NMF), and many others. Also, various similarity measures (cosine, MSD, pearson...) are built-in.

from surprise import Dataset

Is the basic data container in PyMPPA. It serves as the primary form of data storage

from surprise import accuracy

Is the percentage of data that are correctly classified, which ranges from 0 to 1.

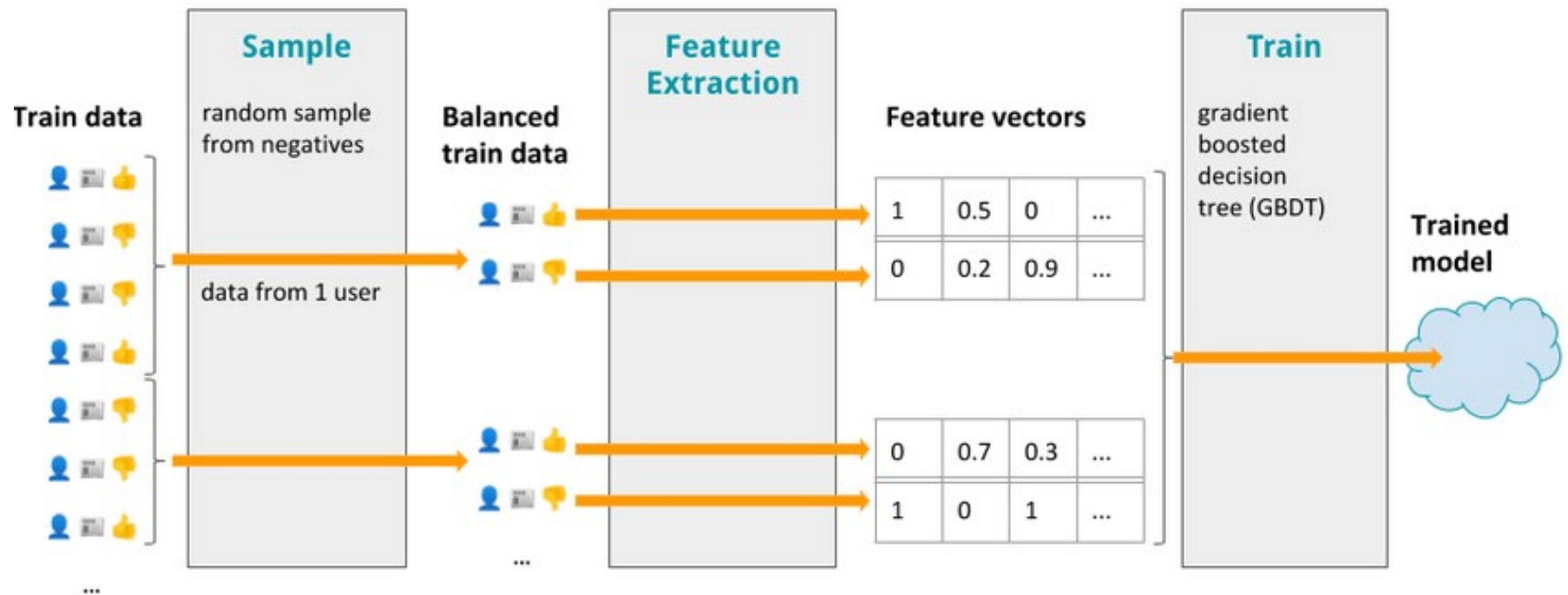


UNIVERSIDAD
SERGIO ARBOLEDA



ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO

How works the pipeline recommenders



What is Docker and why do we use it?

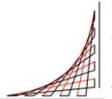


Docker

Docker is a software platform that allows you to build, test, and deploy applications quickly. Docker packages software into standardized units called containers that have everything the software needs to run including libraries, system tools, code, and runtime.

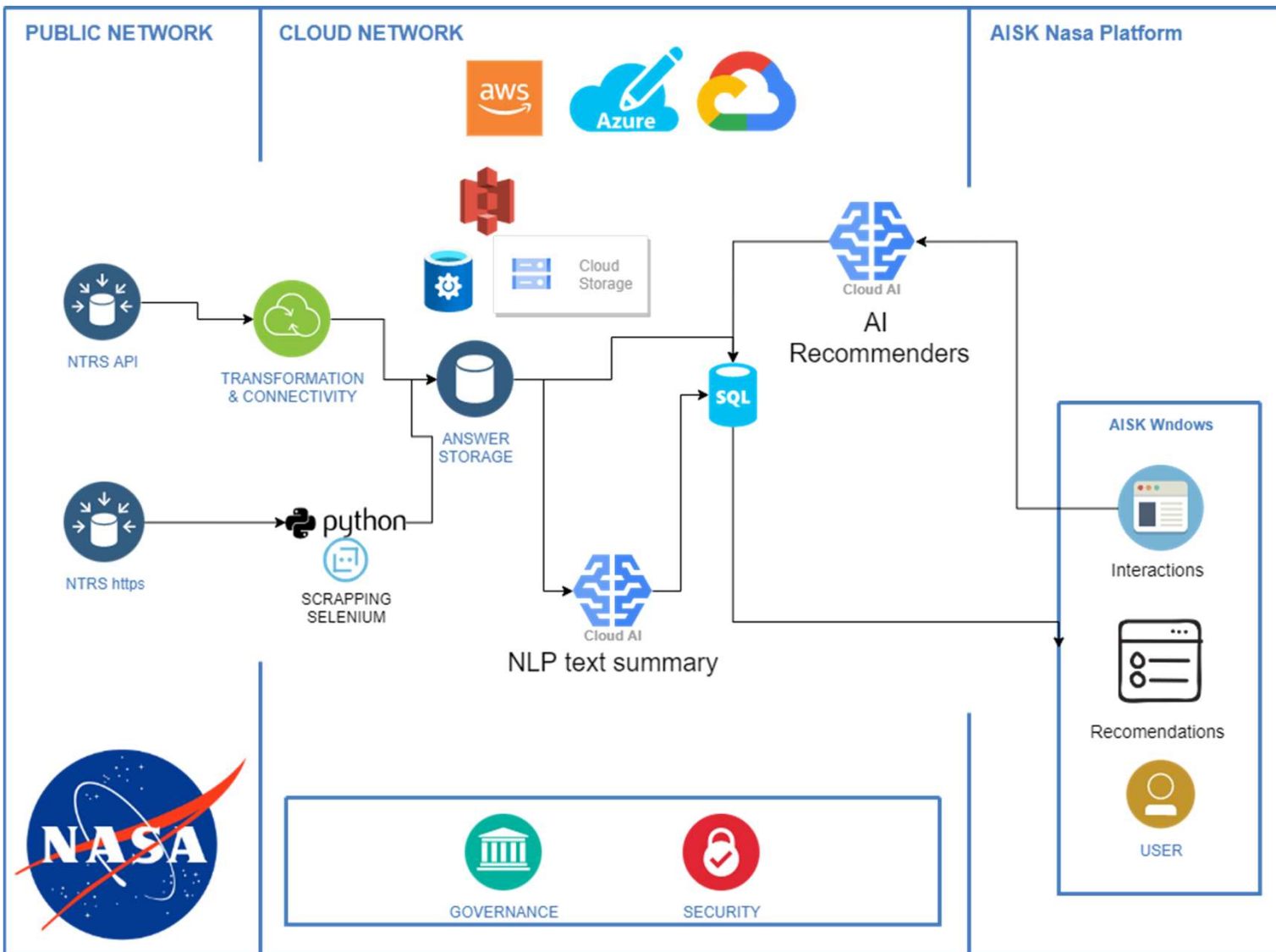


UNIVERSIDAD
SERGIO ARBOLEDA



ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO





AI SK

GitHub Repository

https://github.com/edsteca/AISK_NASA

[edsteca / AISK_NASA](#) Public

[Pin](#) [Unwatch 1](#) [Fork 0](#) [Star 0](#)

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

[main](#) [1 branch](#) [0 tags](#) [Go to file](#) [Add file](#) [Code](#)

edsteca Update README.md 9e0322a 17 seconds ago 24 commits

Architecture	Update readme.md	12 minutes ago
NLP	Create readme.md	1 hour ago
Recommenders	Create readme.md	1 hour ago
Scrap	Create readme.md	1 hour ago
README.md	Update README.md	17 seconds ago

README.md

AIISK_NASA

About

Repository for Spaceappschallenge "Can AI Preserve Our Science Legacy"

[Readme](#)

0 stars

1 watching

0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

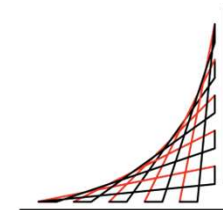
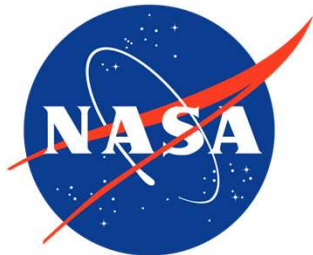


References

Comprehensive Guide to Text Summarization using Deep Learning in Python

<https://www.analyticsvidhya.com/blog/2019/06/comprehensive-guide-text-summarization-using-deep-learning-python/>

- Recommenders <https://github.com/microsoft/recommenders>
- NLP examples <https://github.com/microsoft/nlp-recipes>
- Databases http://www.pymvpa.org/tutorial_datasets.html
- Recomendaciones usando Python <https://ourcodeworld.co/articulos/leer/1101/construyendo-un-sistema-de-recomendacion-simple-en-python>
- Read pdf file using Python <https://www.delftstack.com/es/howto/python/read-pdf-in-python/#:~:text=PDFplumber%20es%20un%20m%C3%B3dulo%20de,para%20leer%20un%20archivo%20PDF.>
- What is S3 <https://aws.amazon.com/es/s3/getting-started/>
- What is a container <https://www.cio.com/article/247005/what-are-containers-and-why-do-you-need-them.html>
- Text Summarizer https://github.com/aravindpai/How-to-build-own-text-summarizer-using-deep-learning/blob/master/How_to_build_own_text_summarizer_using_deep_learning.ipynb

[illegible]

ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO